

*Sub
as
1
2
3*

WHAT IS CLAIMED IS:

1. A method of classifying an instance into one or more classes selected from a set of potential classes, comprising:

3 selecting from the set of potential classes a subset of two or more classes to which the instance is determined to most likely belong; and

5 applying to the instance a scrutiny classifier generated from a set of training records corresponding to a class set inclusive of the selected subset of classes to 6 identify at least one class to which the instance most likely belongs.

1 2. The method of claim 1, wherein the subset of classes is selected based upon assignment to each of the potential classes a probability estimate of the 3 instance belonging to the class.

1 3. The method of claim 2, wherein the selected subset of classes consists 2 of a preselected number of potential classes having highest assigned probability 3 estimates.

1 4. The method of claim 2, wherein the selected subset of classes consists 2 of a number of potential classes having highest assigned probability estimates and a 3 cumulative assigned probability estimate exceeding a preselected threshold.

1 5. The method of claim 2, wherein the probability estimates are assigned 2 to each potential class by applying to the instance a ballpark classifier generated from 3 a set of training records corresponding to the set of potential classes.

1 6. The method of claim 5, wherein the ballpark classifier is generated by a 2 Naïve Bayes inducing algorithm.

1 7. The method of claim 1, wherein the subset of classes is selected based 2 at least in part upon a prescribed misclassification cost.

1 8. The method of claim 1, wherein the scrutiny classifier is generated by a 2 Naïve Bayes inducing algorithm.

3 9. The method of claim 1, wherein the scrutiny classifier is generated by a
4 decision tree inducing algorithm.

1 10. The method of claim 1, further comprising generating the scrutiny
2 classifier from the set of training records.

1 11. The method of claim 10, wherein the scrutiny classifier is generated on-
2 the-fly from a set of training records corresponding to the selected subset of classes.

1 12. The method of claim 10, wherein the scrutiny classifier is generated
2 beforehand in anticipation of the instance to be classified.

1 13. The method of claim 12, wherein the scrutiny classifier is generated
2 based upon an occurrence probability estimate for the inclusive class set.

1 14. The method of claim 13, further comprising selecting an inclusive class
2 set encompassing the selected subset of classes from which to generate the scrutiny
3 classifier.

1 15. The method of claim 1, further comprising applying to the instance a
2 classifier generated from a set of training records corresponding to two or more
3 classes identified by the scrutiny classifier to identify at least one class to which the
4 instance is determined to most likely belong.

1 16. A system for classifying an instance into one or more classes selected
2 from a set of potential classes, comprising:

3 a ballpark classifier configured to select from the set of potential classes a
4 subset of two or more classes to which the instance is determined to most likely
5 belong; and

6 a scrutiny classifier configured to identify from the selected subset of classes
7 at least one class to which the instance most likely belongs.

1 17. The system of claim 16, wherein the ballpark classifier is generated
2 from a set of training records corresponding to the set of potential classes.

1 18. The system of claim 16, wherein the scrutiny classifier is generated
2 from a set of training records corresponding to a class set inclusive of the selected
3 subset of classes.

1 19. The system of claim 18, wherein the scrutiny classifier is generated on-
2 the-fly from a set of training records corresponding to the selected subset of classes.

1 20. A computer program residing on a computer-readable medium for
2 causing a processor executing the computer program to classifying an instance into
3 one or more classes selected from a set of potential classes, the computer program
4 comprising instructions to:

5 select from the set of potential classes a subset of two or more classes to
6 which the instance is determined to most likely belong; and

7 apply to the instance a scrutiny classifier generated from a set of training
8 records corresponding to a class set inclusive of the selected subset of classes to
9 identify at least one class to which the instance most likely belongs.